

Multi-parameter Tikhonov Regularisation in Topological Spaces

Markus Grasmair

Computational Science Center
University of Vienna
Nordbergstr. 15
A-1090 Vienna, Austria

September 1, 2011

Abstract

We study the behaviour of Tikhonov regularisation on topological spaces with multiple regularisation terms. The main result of the paper shows that multi-parameter regularisation is well-posed in the sense that the results depend continuously on the data and converge to a true solution of the equation to be solved as the noise level decreases to zero. Moreover, we derive convergence rates in terms of a generalised Bregman distance using the method of variational inequalities. All the results in the paper, including the convergence rates, consider not only noise in the data, but also errors in the operator.

1 Introduction

Classical Tikhonov regularisation for the approximate solution of an ill-posed operator equation $F(x) = y$ on Hilbert spaces consists in the minimisation of the Tikhonov functional

$$\mathcal{T}(x) := \|F(x) - y\|^2 + \alpha \|x\|^2$$

for some regularisation parameter $\alpha > 0$ depending on the noise level [7, 21]. Here, the regularisation term $\|x\|^2$ encodes some qualitative a-priori knowledge about the true solution of the equation—in this case, it is assumed to have a small Hilbert space norm. In many applications, however, for instance in image processing, the a-priori knowledge has a different form than that of smallness of some Hilbert space norm. Therefore it is necessary to employ other kinds of regularisation terms (see [19] for an overview on regularisation methods in image processing), and one arrives at Tikhonov functionals of the form

$$\mathcal{T}(x) := \|F(x) - y\|^2 + \alpha \mathcal{R}(x)$$

with convex regularisation terms \mathcal{R} . The regularising properties of Tikhonov functionals of that form have for instance been studied in [20].

In this paper, we study two additional generalisations of Tikhonov regularisation. First, we consider more general, non-quadratic, distance like measures for the similarity term instead of the squared norm of the residual. Typical examples include *f-divergences*, which appear naturally when the noise is known to follow a distribution that is not Gaussian. An overview of useful similarity terms for Tikhonov regularisation can for instance be found in [18]. The second generalisation is concerned with the regularisation term. Instead of assuming that the a-priori information about the true solution can be encoded in a single functional, we study the situation where we have available different, possibly contradicting pieces of information, each of which can be described by the smallness of a different regularisation term \mathcal{R}_k . Examples in image processing include the famous Mumford–Shah model, which assumes that images consist of different objects, characterised by smooth intensity variations and separated by pronounced edges [17], but also models, which decompose images into geometric parts and texture [16, 22] (see also [1] for an overview). Other examples, where multi-parameter Tikhonov regularisation has been proposed for the solution of inverse problems include [3, 4, 6, 23]. In all these settings, regularisation is achieved by minimisation of a Tikhonov functional of the form

$$\mathcal{T}(x) := \mathcal{S}(F(x), y) + \sum_k \alpha_k \mathcal{R}_k(x). \quad (1)$$

From the theoretical point of view, multi-parameter Tikhonov regularisation is interesting, as it shares most of the features of single-parameter Tikhonov regularisation, but still exhibits some crucial differences, for instance concerning the formulation and interpretation of convergence rates. Still, it seems that no comprehensive study concerning the regularising properties of multi-parameter Tikhonov regularisation has been published. Most of the existing theoretical papers rather deal primarily with the problem of a suitable parameter choice (see for instance [13, 14, 15]), but questions like stability and convergence have been neglected, in particular in the interesting case when the different regularisation parameters α_k decrease to zero at different rates.

In this paper, we will prove stability and convergence of multi-parameter regularisation under fairly general conditions. Because anyway no trace of an original Hilbert space or Banach space structure is left in the formulation of the Tikhonov functional \mathcal{T} in (1), we will completely discard all assumption of a linear structure and instead consider the situation, where both the domain X and the co-domain Y of the operator F are mere topological spaces, with the topology of Y defined by the distance measure \mathcal{S} . In this setting, we prove the continuous dependence of the minimiser $x \in X$ of \mathcal{T} on variations in the data $y \in Y$ and the regularisation vector α , as long as at least one component of α stays positive. In addition, we consider the case where also the operator F to be inverted is prone to errors. We introduce a suitable topology on the space of all operators from X to Y , which is related to the topology of uniform convergence on compact sets, and show that the minimisers of \mathcal{T} also depend continuously on the operator F .

Then we turn to the study of the behaviour of the minimisers of \mathcal{T} as the regularisation vector and the noise level—both noise in the data and the operator—tend to zero. We prove the convergence of the regularised solutions to a solution of the exact equation provided the regularisation vector converges to zero sufficiently slowly. The relation between the noise level and the regularisation that

is required for deriving this kind of convergence is a natural generalisation of the usual condition required for the convergence of single-parameter Tikhonov regularisation. There is, however, a notable difference to the single-parameter case. While in the single-parameter case, the regularised solutions converge not to any solution of the equation $F(x) = y$, but rather to an \mathcal{R} -minimising one, this behaviour cannot be guaranteed in the multi-parameter case, if the components of α decrease to zero at different rates and the regularisation terms \mathcal{R}_k have different proper domains.

Finally, we derive quantitative estimates for the difference between the regularised solution of the equation and the true solution in dependence of the noise level, the regularisation parameter, and the accuracy of the operator. Because we work in general topological spaces, we cannot employ the classical range or source conditions for the derivation of convergence rates. Instead, we make use of the method of variational inequalities introduced in [12] and its modifications and generalisations used in [2, 10]. Apart from the extension to multi-parameter regularisation, a major novelty of the quantitative estimate lies in the inclusion of operator errors, which previously have never been treated by means of variational inequalities.

2 Preliminaries

Assume that X and Y are sets and $F: X \rightarrow Y$ some mapping. We consider multi-parameter Tikhonov regularisation with a regularisation functional of the form

$$\mathcal{T}(x; \alpha, y, F) = \mathcal{S}(F(x), y) + \sum_k \alpha_k \mathcal{R}_k(x), \quad (2)$$

where $\mathcal{R}_k: X \rightarrow [0, +\infty]$ are non-negative regularising terms and $\mathcal{S}: Y \times Y \rightarrow \mathbb{R}_{\geq 0}$ is a distance like functional satisfying $\mathcal{S}(y, z) = 0$ if and only if $y = z$. Here we define $\alpha_k \mathcal{R}_k(x) := 0$ if $\alpha_k = 0$ and $\mathcal{R}_k(x) = +\infty$.

Single-parameter Tikhonov regularisation in such a general setting with non-metric distance measure has been considered in [8, 9, 18]. See in particular [18], where a large number of useful similarity measures is presented. In all these papers it was assumed that the target space Y is a topological space with a topology that is well compatible with both the function F to be inverted and the similarity measure \mathcal{S} . In this paper, we follow the approach from [11], where the well-posedness of the residual method has been treated, and use the similarity measure \mathcal{S} to define a topology on Y .

First, we consider on the set Y the uniformity \mathcal{U} that is induced by the family of pseudo-metrics $d^{(z)}: Y \times Y \rightarrow \mathbb{R}_{\geq 0}$, $z \in Y$, defined by

$$d^{(z)}(y, \tilde{y}) := |\mathcal{S}(z, y) - \mathcal{S}(z, \tilde{y})|.$$

Moreover, we denote by σ the topology induced by the uniformity \mathcal{U} . Then a sequence $\{y^{(l)}\}_{l \in \mathbb{N}} \subset Y$ converges to $y \in Y$ with respect to σ , if $\mathcal{S}(z, y^{(l)}) \rightarrow \mathcal{S}(z, y)$ for every $z \in Y$. Note in particular that the condition $\mathcal{S}(z, y) = 0$ if and only if $y = z$ implies that the topology σ is Hausdorff, as the metric $d^{(z)}$ separates z from every other point $y \in Y \setminus \{z\}$.

Remark 2.1. Consider the special case where the functional \mathcal{S} is of the form $\mathcal{S} = \rho \circ d$ with d a metric on Y and $\rho: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ continuous and strictly increasing

with $\rho(0) = 0$. Because for every $z \in Y$ the mapping $y \mapsto \mathcal{S}(z, y) = \rho(d(z, y))$ is continuous with respect to the metric topology on Y , it follows that the metric topology is finer than σ . Conversely, every metric ball

$$B_r(z) = \{y \in Y : d(z, y) < r\} = \{y \in Y : \mathcal{S}(z, y) < \rho^{-1}(r)\}$$

is open with respect to σ , which shows that, in fact, the topology σ coincides with the metric topology on Y . \blacksquare

Remark 2.2. In [11], a topology on Y has been defined by the single pseudo-metric

$$d^{\text{sup}}(y, \tilde{y}) := \sup_{z \in Y} d^{(z)}(y, \tilde{y}) = \sup \left\{ |\mathcal{S}(z, y) - \mathcal{S}(z, \tilde{y})| : z \in Y \right\}. \quad (3)$$

While this definition is reasonable in the case of the residual method, where we can assume that the functional \mathcal{S} has a structure that is very similar to that of a distance, it is less so for Tikhonov regularisation, where we must rather assume that \mathcal{S} resembles the *power* of a distance. Indeed, already in the classical Hilbert space setting with $\mathcal{S}(z, y) = \|z - y\|_Y^2$ the definition (3) is useless, as $d^{\text{sup}}(y, \tilde{y}) = +\infty$ whenever $y \neq \tilde{y}$. \blacksquare

In addition to the topology on Y , we require a suitable topology on the space X , which ensures the existence of a minimiser of the regularisation functional $\mathcal{T}(\cdot; \alpha, y, F)$ for every positive regularisation parameter $\alpha \in \mathbb{R}_{\geq 0}^n \setminus \{0\}$, every $y \in Y$, and all mappings F that are compatible with the topologies on X and Y . The first assumption on the topology is a standard requirement for the subsequent application of the direct method in the calculus of variations.

Assumption 2.3. There exists a topology τ on X such that each mapping $\mathcal{R}_k : X \rightarrow [0, +\infty]$ is sequentially coercive and lower semi-continuous with respect to τ . \blacksquare

In addition, we have to ensure that the mapping F is well compatible with the topology τ on X . More precisely, we require the lower semi-continuity of the mapping $(x, y) \mapsto \mathcal{S}(F(x), y)$. This condition alone, however, is not sufficient for obtaining stability when one of the regularisation parameters tends to zero (but not all of them do). In order to treat this case as well, we have to introduce a density condition for the domains of the regularisation terms \mathcal{R}_k with respect to a suitable type of convergence.

We denote by

$$\mathcal{D} := \{x \in X : \mathcal{R}_k(x) < +\infty \text{ for all } 1 \leq k \leq n\}$$

the joint domain of the regularisation terms \mathcal{R}_k . Throughout the paper, we assume that $\mathcal{D} \neq \emptyset$; in the degenerate case $\mathcal{D} = \emptyset$, multi-parameter regularisation with the regularisation terms \mathcal{R}_k is not very useful, as, necessarily, some of the regularisation parameters α_k have to be zero.

Definition 2.4. Let τ be a topology on X . We denote by $\mathcal{F}(\tau)$ the set of all mappings $F : X \rightarrow Y$ satisfying the following conditions:

- The mapping $(x, y) \mapsto \mathcal{S}(F(x), y)$ is sequentially lower semi-continuous with respect to the product topology $(\tau \times \sigma)$ on $X \times Y$.

- For every $x \in X$ and every $y \in Y$ there exists a sequence $\{x^{(l)}\}_{l \in \mathbb{N}} \subset \mathcal{D}$ converging to x with respect to τ such that $\mathcal{R}_k(x^{(l)}) \rightarrow \mathcal{R}_k(x)$ for every $1 \leq k \leq n$ and $\mathcal{S}(F(x^{(l)}), y) \rightarrow \mathcal{S}(F(x), y)$. \blacksquare

Because we study also the stability of Tikhonov regularisation with respect to operator errors, we have to introduce in addition some notion of convergence of operators $F: X \rightarrow Y$. To that end, we define for $K \subset X$ and $L \subset Y$ a pseudo-metric $d_{K,L}$ on the set of all functions from X to Y by

$$d_{K,L}(F, G) := \sup\{|\mathcal{S}(F(x), z) - \mathcal{S}(G(x), z)| : x \in K, z \in L\}.$$

Definition 2.5. Let $F^{(l)}: X \rightarrow Y$, $l \in \mathbb{N}$, be a sequence of mappings, let $F: X \rightarrow Y$, and let τ be a topology on X . We say that the sequence $\{F^{(l)}\}_{l \in \mathbb{N}}$ converges to F with respect to τ , if

$$d_{K,L}(F^{(l)}, F) \rightarrow 0$$

whenever $K \subset X$ is sequentially τ -compact and $L \subset Y$ is sequentially σ -compact. \blacksquare

Remark 2.6. Note that, according to Definition 2.5, a sequence of functions $F^{(l)}: X \rightarrow Y$ convergence to $F: X \rightarrow Y$ with respect to τ , if and only if the sequence of real valued functions $(x, y) \mapsto \mathcal{S}(F^{(l)}(x), y)$ converges to the function $(x, y) \mapsto \mathcal{S}(F(x), y)$ uniformly on sequentially compact sets. \blacksquare

The following implications of the convergence of functions introduced in 2.5 will be required several times in this paper.

Lemma 2.7. Assume that the sequence of functions $\{F^{(l)}\}_{l \in \mathbb{N}} \subset \mathcal{F}(\tau)$ converges to $F \in \mathcal{F}(\tau)$ with respect to τ . Assume moreover that $\{x^{(l)}\}_{l \in \mathbb{N}} \subset X$ converges to $x \in X$ with respect to τ and $\{y^{(l)}\}_{l \in \mathbb{N}} \subset Y$ converges to $y \in Y$ with respect to σ . Then

$$\lim_l \mathcal{S}(F^{(l)}(x), y^{(l)}) = \mathcal{S}(F(x), y) \leq \liminf_l \mathcal{S}(F^{(l)}(x^{(l)}), y^{(l)}).$$

Proof. Consider the sequentially compact sets $K := \{x\} \cup \{x^{(l)}\}_{l \in \mathbb{N}}$ and $L := \{y\} \cup \{y^{(l)}\}_{l \in \mathbb{N}}$. Then

$$\begin{aligned} & |\mathcal{S}(F^{(l)}(x), y^{(l)}) - \mathcal{S}(F(x), y)| \\ & \leq |\mathcal{S}(F^{(l)}(x), y^{(l)}) - \mathcal{S}(F(x), y^{(l)})| + |\mathcal{S}(F(x), y^{(l)}) - \mathcal{S}(F(x), y)| \\ & \leq d_{x,L}(F^{(l)}, F) + |\mathcal{S}(F(x), y^{(l)}) - \mathcal{S}(F(x), y)|. \end{aligned}$$

Now the convergence of the sequence $\{F^{(l)}\}_{l \in \mathbb{N}}$ to F implies that $d_{x,L}(F^{(l)}, F) \rightarrow 0$. In addition, the convergence $y^{(l)} \rightarrow y$ implies that $\mathcal{S}(F(x), y^{(l)}) \rightarrow \mathcal{S}(F(x), y)$. This shows that

$$\lim_l \mathcal{S}(F^{(l)}(x), y^{(l)}) = \mathcal{S}(F(x), y).$$

Moreover we have

$$\begin{aligned} \mathcal{S}(F^{(l)}(x^{(l)}), y^{(l)}) &= \mathcal{S}(F(x^{(l)}), y^{(l)}) + \mathcal{S}(F^{(l)}(x^{(l)}), y^{(l)}) - \mathcal{S}(F(x^{(l)}), y^{(l)}) \\ &\geq \mathcal{S}(F(x^{(l)}), y^{(l)}) - d_{K,L}(F^{(l)}, F). \end{aligned}$$

Thus the convergence of $\{F^{(l)}\}_{l \in \mathbb{N}}$ to F and the fact that $F \in \mathcal{F}(\tau)$ imply that

$$\begin{aligned} \liminf_l \mathcal{S}(F^{(l)}(x^{(l)}), y^{(l)}) &\geq \liminf_l \mathcal{S}(F(x^{(l)}), y^{(l)}) - \lim_l d_{K,L}(F^{(l)}, F) \\ &\geq \mathcal{S}(F(x), y) . \end{aligned} \quad \square$$

3 Well-posedness

In this section we study the well-posedness of multi-parameter Tikhonov regularisation with the Tikhonov functional \mathcal{T} given in (2). First we prove the existence of a minimiser, and we show that the minimisers depend continuously on the data $y \in Y$, the regularisation parameter $\alpha \in \mathbb{R}_{\geq 0}^n$, and the operator $F: X \rightarrow Y$. Then we prove the convergence of the minimisers to a solution of the equation $F(x) = y$ as the noise level approaches zero, as long as the regularisation parameters α_k tend to zero sufficiently slowly.

3.1 Existence and Stability

Proposition 3.1. *Let Assumption 2.3 be satisfied, let $F \in \mathcal{F}(\tau)$, $y \in Y$, and $\alpha \in \mathbb{R}_{\geq 0}^n \setminus \{0\}$. Then the Tikhonov functional $\mathcal{T}(\cdot; \alpha, y, F)$ admits a minimum in X .*

Proof. This is a straightforward application of the direct method in the calculus of variations. \square

The main difficulties in the proof of the stability of the regularisation method are due to the incorporation of operator errors, but also due to the fact that we do not exclude the situation, where some of the regularisation parameters α_k vanish (though we require that at least one of them stays positive).

Proposition 3.2. *Let Assumption 2.3 be satisfied, let $\{y^{(l)}\}_{l \in \mathbb{N}} \subset Y$ be any sequence converging to $y^\delta \in Y$ with respect to σ , let $\{F^{(l)}\}_{l \in \mathbb{N}} \subset \mathcal{F}(\tau)$ be any sequence converging to $F^\delta \in \mathcal{F}(\tau)$, and let $\{\alpha^{(l)}\}_{l \in \mathbb{N}} \subset \mathbb{R}_{\geq 0}^n \setminus \{0\}$ converge to $\alpha \in \mathbb{R}_{\geq 0}^n \setminus \{0\}$, and let*

$$x^{(l)} \in \arg \min \{ \mathcal{T}(x; \alpha^{(l)}, y^{(l)}, F^{(l)}) : x \in X \} .$$

Then the sequence $\{x^{(l)}\}_{l \in \mathbb{N}}$ has a sub-sequence that converges with respect to τ to some

$$x_\alpha \in \arg \min \{ \mathcal{T}(x; \alpha, y^\delta, F^\delta) : x \in X \} .$$

Proof. Let $\tilde{x} \in \mathcal{D}$ be arbitrary and let $1 \leq k_0 \leq n$ be such that $\alpha_{k_0} > 0$. By the minimality of $x^{(l)}$ we have

$$\begin{aligned} \alpha_{k_0}^{(l)} \mathcal{R}_{k_0}(x^{(l)}) &\leq \mathcal{T}(x^{(l)}; \alpha^{(l)}, y^{(l)}, F^{(l)}) \\ &\leq \mathcal{T}(\tilde{x}; \alpha^{(l)}, y^{(l)}, F^{(l)}) \\ &= \mathcal{S}(F^{(l)}(\tilde{x}), y^{(l)}) + \sum_k \alpha_k^{(l)} \mathcal{R}_k(\tilde{x}) . \end{aligned}$$

Because the sequence $\{\alpha^{(l)}\}_{l \in \mathbb{N}}$ converges to α and $\alpha_{k_0} > 0$, we may assume without loss of generality that $\alpha_{k_0}^{(l)} > 0$ for all $l \in \mathbb{N}$, and therefore also $\inf_l \alpha_{k_0}^{(l)} > 0$. In addition, it follows that $\sup_l \alpha_k^{(l)} < +\infty$ for every k . Moreover, the convergence of the sequence $\{y^{(l)}\}_{l \in \mathbb{N}}$ to y^δ implies that $\mathcal{S}(F^{(l)}(\tilde{x}), y^{(l)}) \rightarrow \mathcal{S}(F(\tilde{x}), y^\delta) < +\infty$ (see Lemma 2.7), showing that $\sup_l \mathcal{S}(F^{(l)}(\tilde{x}), y^{(l)}) < +\infty$. Therefore

$$\sup_l \mathcal{R}_{k_0}(x^{(l)}) \leq \sup_l \frac{\mathcal{S}(F^{(l)}(\tilde{x}), y^{(l)}) + \sum_k \alpha_k^{(l)} \mathcal{R}_k(\tilde{x})}{\alpha_{k_0}^{(l)}} < +\infty.$$

The sequential τ -coercivity of \mathcal{R}_{k_0} implies now the existence of a sub-sequence, for simplicity again denoted by $\{x^{(l)}\}_{l \in \mathbb{N}}$, converging with respect to τ to some $x_0 \in X$.

It remains to show that $\mathcal{T}(x_0; \alpha, y^\delta, F^\delta) \leq \mathcal{T}(x; \alpha, y^\delta, F^\delta)$ for every $x \in X$. To that end, note first that Lemma 2.7 implies that

$$\mathcal{S}(F^\delta(x_0), y^\delta) \leq \liminf_l \mathcal{S}(F^{(l)}(x^{(l)}), y^{(l)}).$$

Because the functionals \mathcal{R}_k are τ -lower semi-continuous and $\alpha^{(l)} \rightarrow \alpha$, it follows that also

$$\sum_k \alpha_k \mathcal{R}_k(x_0) \leq \liminf_l \sum_k \alpha_k^{(l)} \mathcal{R}_k(x^{(l)}).$$

Thus

$$\mathcal{T}(x_0; \alpha, y^\delta, F^\delta) \leq \liminf_l \mathcal{T}(x^{(l)}; \alpha^{(l)}, y^{(l)}, F^{(l)}). \quad (4)$$

Now assume that $\tilde{x} \in \mathcal{D}$ is arbitrary. Then we obtain from Lemma 2.7 the equality

$$\mathcal{S}(F^\delta(\tilde{x}), y^\delta) = \lim_l \mathcal{S}(F^{(l)}(\tilde{x}), y^{(l)}).$$

Moreover, since $\mathcal{R}_k(\tilde{x}) < +\infty$ for every k , it follows that

$$\sum_k \alpha_k^{(l)} \mathcal{R}_k(\tilde{x}) \rightarrow \sum_k \alpha_k \mathcal{R}_k(\tilde{x}).$$

Thus (4) and the minimality assumption of $x^{(l)}$ imply that

$$\begin{aligned} \mathcal{T}(x_0; \alpha, y^\delta, F^\delta) &\leq \liminf_l \mathcal{T}(x^{(l)}; \alpha^{(l)}, y^{(l)}, F^{(l)}) \\ &\leq \liminf_l \mathcal{T}(\tilde{x}; \alpha^{(l)}, y^{(l)}, F^{(l)}) \\ &= \liminf_l \left(\mathcal{S}(F^{(l)}(\tilde{x}), y^{(l)}) + \sum_k \alpha_k^{(l)} \mathcal{R}_k(\tilde{x}) \right) \\ &= \mathcal{S}(F^\delta(\tilde{x}), y^\delta) + \sum_k \alpha_k \mathcal{R}_k(\tilde{x}) \\ &= \mathcal{T}(\tilde{x}; \alpha, y^\delta, F^\delta). \end{aligned} \quad (5)$$

Now let $\tilde{x} \in X$ be arbitrary. Then there exists a sequence $\{\tilde{x}^{(l)}\}_{l \in \mathbb{N}}$ converging to \tilde{x} with respect to τ such that $\tilde{x}^{(l)} \in \mathcal{D}$ for every l , $\mathcal{R}_k(\tilde{x}^{(l)}) \rightarrow \mathcal{R}_k(\tilde{x})$ for every $1 \leq k \leq n$, and $\mathcal{S}(F^\delta(\tilde{x}^{(l)}), y^\delta) \rightarrow \mathcal{S}(F^\delta(\tilde{x}), y^\delta)$. Consequently (5) implies that

$$\mathcal{T}(x_0; \alpha, y^\delta, F^\delta) \leq \liminf_l \mathcal{T}(\tilde{x}^{(l)}; \alpha, y^\delta, F^\delta) = \mathcal{T}(\tilde{x}; \alpha, y^\delta, F^\delta), \quad (6)$$

showing that x_0 is indeed a minimiser of $\mathcal{T}(\cdot; \alpha, y^\delta, F^\delta)$. \square

3.2 Convergence

In standard Tikhonov regularisation, the usual condition guaranteeing convergence is the assumption $\delta^2/\alpha \rightarrow 0$. In the setting of multi-parameter regularisation, this assumption obviously cannot be applied directly. One possible remedy are the assumptions

$$\frac{\min_k \alpha_k}{\max_k \alpha_k} > c_0 > 0 \quad \text{and} \quad \frac{\mathcal{S}(y, y^\delta)}{\max_k \alpha_k} \rightarrow 0. \quad (7)$$

These assumptions, however, imply that all the regularisation parameters converge to zero at the same speed, which does not seem reasonable. One of the main advantages of multi-parameter regularisation is precisely its flexibility in the parameter choice, which allows for different rates for the different components of the parameter vector. Instead of using (7), we therefore consider, in the case of an exact operator F , the weaker condition

$$\frac{\mathcal{S}(y, y^\delta)}{\sum_k \alpha_k} \rightarrow 0.$$

This condition allows the different parameters to decrease at different rates. Moreover, it makes sense, as one can interpret the denominator, $\sum_k \alpha_k$, as the total amount of regularisation.

Theorem 3.3. *Let Assumption 2.3 be satisfied, let $F \in \mathcal{F}(\tau)$ and $y \in Y$, and assume that there exists $x_0 \in \mathcal{D}$ such that $F(x_0) = y$. Let moreover $\{y^{(l)}\}_{l \in \mathbb{N}} \subset Y$ be any sequence converging to y with respect to σ , and $\{F^{(l)}\}_{l \in \mathbb{N}} \subset \mathcal{F}(\tau)$ any sequence converging to F with respect to τ . Let $L := \{y\} \cup \{y^{(l)}\}_{l \in \mathbb{N}}$ and*

$$K := \{x \in X : \mathcal{R}_k(x) \leq \mathcal{R}_k(x_0) + 1 \text{ for some } 1 \leq k \leq n\}.$$

Assume that

$$\frac{\mathcal{S}(y, y^{(l)}) + d_{K,L}(F^{(l)}, F)}{\sum_k \alpha_k^{(l)}} \rightarrow 0 \quad \text{and} \quad \sum_k \alpha_k^{(l)} \rightarrow 0 \quad (8)$$

as $l \rightarrow \infty$ and let

$$x^{(l)} \in \arg \min \{ \mathcal{T}(x; \alpha^{(l)}, y^{(l)}, F^{(l)}) : x \in X \}.$$

Consider any sub-sequence, again indexed by l , for which the vectors

$$\bar{\alpha}^{(l)} := \frac{\alpha^{(l)}}{\sum_k \alpha_k^{(l)}} \in [0, 1]^n$$

converge to some $\bar{\alpha} \in [0, 1]^n$ (such a sub-sequence exists because of the boundedness of the sequence). Then every sub-sequence of $\{x^{(l)}\}_{l \in \mathbb{N}}$ has a sub-sequence, for simplicity again denoted by $\{x^{(l)}\}_{l \in \mathbb{N}}$, τ -converging to some $x^\dagger \in X$ satisfying $F(x^\dagger) = y$. In addition

$$\sum_k \bar{\alpha}_k \mathcal{R}_k(x^\dagger) \leq \inf \left\{ \sum_k \bar{\alpha}_k \mathcal{R}_k(x) : x \in \mathcal{D}, F(x) = y \right\}. \quad (9)$$

Proof. Because the functionals \mathcal{R}_k are sequentially τ -coercive, it follows that the set K is sequentially τ -compact as the finite union of sequentially τ -compact sets. Moreover, the definition of $x^{(l)}$ implies that

$$\begin{aligned} \sum_k \alpha_k^{(l)} \mathcal{R}_k(x^{(l)}) &\leq \mathcal{T}(x^{(l)}; \alpha^{(l)}, y^{(l)}, F^{(l)}) \\ &\leq \mathcal{T}(x_0; \alpha^{(l)}, y^{(l)}, F^{(l)}) \\ &= \mathcal{S}(F^{(l)}(x_0), y^{(l)}) + \sum_k \alpha_k^{(l)} \mathcal{R}_k(x_0) \\ &\leq \mathcal{S}(F(x_0), y^{(l)}) + d_{K,L}(F^{(l)}, F) + \sum_k \alpha_k^{(l)} \mathcal{R}_k(x_0). \end{aligned}$$

Dividing by $\sum_k \alpha_k^{(l)}$ and using (8) and the facts that $F(x_0) = y$ and $x_0 \in \mathcal{D}$, we see that

$$\limsup_l \sum_k \bar{\alpha}_k^{(l)} \mathcal{R}_k(x^{(l)}) \leq \limsup_l \sum_k \bar{\alpha}_k^{(l)} \mathcal{R}_k(x_0) = \sum_k \bar{\alpha}_k \mathcal{R}_k(x_0) < +\infty.$$

Thus also

$$\limsup_l \sum_k \bar{\alpha}_k \mathcal{R}_k(x^{(l)}) \leq \limsup_l \sum_k \bar{\alpha}_k^{(l)} \mathcal{R}_k(x^{(l)}) \leq \sum_k \bar{\alpha}_k \mathcal{R}_k(x_0). \quad (10)$$

In particular, for l sufficiently large,

$$\sum_k \bar{\alpha}_k \mathcal{R}_k(x^{(l)}) \leq \sum_k \bar{\alpha}_k \mathcal{R}_k(\tilde{x}) + 1 = \sum_k \bar{\alpha}_k (\mathcal{R}_k(\tilde{x}) + 1),$$

which in turn proves that the sequence $\{x^{(l)}\}_{l \in \mathbb{N}}$ is eventually contained in the sequentially τ -compact set K . Thus, there exists a sub-sequence, for simplicity again denoted by $\{x^{(l)}\}_{l \in \mathbb{N}}$, converging with respect to τ to some $x^\dagger \in K$.

Now note that Lemma 2.7 and the facts that $x^{(l)} \rightarrow x^\dagger$, $y^{(l)} \rightarrow y$, $F^{(l)} \rightarrow F$, and $\sum_k \alpha_k^{(l)} \rightarrow 0$ imply that

$$\begin{aligned} \mathcal{S}(F(x^\dagger), y) &\leq \liminf_l \mathcal{S}(F^{(l)}(x^{(l)}), y^{(l)}) \\ &\leq \liminf_l \mathcal{T}(x^{(l)}; \alpha^{(l)}, y^{(l)}, F^{(l)}) \\ &\leq \liminf_l \mathcal{T}(x_0; \alpha^{(l)}, y^{(l)}, F^{(l)}) \\ &\leq \liminf_l \left(\mathcal{S}(y, y^{(l)}) + d_{K,L}(F^{(l)}, F) + \sum_k \alpha_k^{(l)} \mathcal{R}_k(x_0) \right) \\ &= 0, \end{aligned}$$

showing that $\mathcal{S}(F(x^\dagger), y) = 0$ and hence $F(x^\dagger) = y$. Moreover, the τ -lower semi-continuity of the functions \mathcal{R}_k and the fact that (10) holds for every $\tilde{x} \in \mathcal{D} \cap K$ satisfying $F(\tilde{x}) = y$ now imply that

$$\begin{aligned} \sum_k \bar{\alpha}_k \mathcal{R}_k(x^\dagger) &\leq \liminf_l \sum_k \bar{\alpha}_k \mathcal{R}_k(x^{(l)}) \leq \liminf_l \sum_k \bar{\alpha}_k^{(l)} \mathcal{R}_k(x^{(l)}) \\ &\leq \inf \left\{ \sum_k \bar{\alpha}_k \mathcal{R}_k(\tilde{x}) : \tilde{x} \in \mathcal{D} \cap K, F(\tilde{x}) = y \right\}. \quad (11) \end{aligned}$$

Now the definition of K implies that also (9) holds, which concludes the proof. \square

Note that there is a crucial difference between the convergence result of Theorem 3.3 and the convergence results that can be derived for single-parameter regularisation. There, using conditions analogous to those of Theorem 3.3, one can show that the limit x^\dagger is an \mathcal{R} -minimising solution of the equation $F(x) = y$. The corresponding result for multi-parameter regularisation would be that x^\dagger is a $\sum_k \bar{\alpha}_k \mathcal{R}_k$ minimising solution of the equation $F(x) = y$, that is,

$$x^\dagger \in \arg \min \left\{ \sum_k \bar{\alpha}_k \mathcal{R}_k(x) : x \in X, F(x) = y \right\}.$$

This assertion, however, need not be true in the case where one of the limiting regularisation parameters $\bar{\alpha}_k$ vanishes, but the approaching weighted parameters $\bar{\alpha}_k^{(l)}$ are all positive.

4 Convergence Rates

For the derivation of convergence rates, or, rather, quantitative estimates for the distance between regularised and true solution, we apply the method of variational inequalities, which has been introduced in [12] (see also [19]) and further developed in [2, 9, 10]. In a Banach space setting with convex regularisation terms, estimates have been classically derived with respect to the Bregman distance, which measures the distance between the regularisation term \mathcal{R}_k and its affine approximation at the true solution x^\dagger (see [5]). Here we consider the setting introduced in [10] (see also [11]) and assume that a variational inequality is satisfied with respect to any distance like functional, which at the same time serves as the distance with respect to which the convergence rates are derived.

Assumption 4.1. The element $x^\dagger \in \mathcal{D}$ satisfies $F(x^\dagger) = y$, and for every $1 \leq k \leq n$ there exists a function $D^k(\cdot; x^\dagger): X \rightarrow \mathbb{R}_{\geq 0}$ satisfying $D^k(x^\dagger; x^\dagger) = 0$ and a concave and strictly increasing function $\Phi_k: \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ with $\Phi_k(0) = 0$ such that

$$D^k(x; x^\dagger) \leq \mathcal{R}_k(x) - \mathcal{R}_k(x^\dagger) + \Phi_k(\mathcal{S}(F(x), y)) \quad (12)$$

for every $x \in X$. ■

In addition to the variational inequalities (12), we assume in this section that the functional \mathcal{S} satisfies a quasi-triangle inequality of the form

$$\mathcal{S}(z_1, z_2) \leq s(\mathcal{S}(z_1, z_3) + \mathcal{S}(z_3, z_2)) \quad (13)$$

for some $s \geq 1$ and every $z_1, z_2, z_3 \in Y$. While such a quasi-triangle inequality is satisfied in the important case where the distance measure \mathcal{S} is the power of some metric on Y , it need not hold for instance in the case where \mathcal{S} is some Bregman distance.

Theorem 4.2. *Let Assumption 4.1 be satisfied and assume that \mathcal{S} satisfies the quasi-triangle inequality (13). Define the function $\Psi: \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}_{\geq 0}$,*

$$\Psi(\alpha) := \sup_{t > 0} \left(\sum_k s^2 \alpha_k \Phi_k(t) - t \right),$$

and let

$$K := \{x \in X : \mathcal{R}_k(x) \leq \mathcal{R}_k(x^\dagger) + 1 \text{ for some } 1 \leq k \leq n\}.$$

Let $y^\delta \in Y$ and $F^\delta \in \mathcal{F}(\tau)$ and

$$x_\alpha^\delta \in \arg \min \{\mathcal{T}(x; \alpha, y^\delta, F^\delta) : x \in X\}.$$

Then the inequality

$$\sum_k \alpha_k D^k(x_\alpha^\delta; x^\dagger) \leq s d_{K,y}(F^\delta, F) + (s+1)\mathcal{S}(y, y^\delta) + \frac{1}{s}\mathcal{S}(y^\delta, y) + \frac{\Psi(\alpha)}{s^2}$$

holds for $d_{K,y}(F^\delta, F)$ and $\mathcal{S}(y, y^\delta)$ sufficiently small. In particular

$$D^j(x_\alpha^\delta; x^\dagger) \leq \frac{s^3 d_{K,y}(F^\delta, F) + (s^3 + s)\mathcal{S}(y, y^\delta) + s\mathcal{S}(y^\delta, y) + \Psi(\alpha)}{s^2 \alpha_j} \quad (14)$$

for every $1 \leq j \leq n$ and every $\alpha \in \mathbb{R}_{\geq 0}^n$ with $\alpha_j > 0$.

Proof. Because x_α^δ is a minimiser of $\mathcal{T}(\cdot; \alpha, y^\delta, F^\delta)$, the inequality

$$\mathcal{S}(F^\delta(x_\alpha^\delta), y^\delta) + \sum_k \alpha_k \mathcal{R}_k(x_\alpha^\delta) \leq \mathcal{S}(F^\delta(x^\dagger), y^\delta) + \sum_k \alpha_k \mathcal{R}_k(x^\dagger) \quad (15)$$

holds. Moreover the quasi-triangle inequality (13) and the fact that $F(x^\dagger) = y$ and hence $\mathcal{S}(F(x^\dagger), y) = 0$ imply that

$$\mathcal{S}(F^\delta(x^\dagger), y^\delta) \leq s(\mathcal{S}(F^\delta(x^\dagger), y) + \mathcal{S}(y, y^\delta)) \leq s(d_{K,y}(F^\delta, F) + \mathcal{S}(y, y^\delta))$$

and

$$\mathcal{S}(F(x_\alpha^\delta), y^\delta) \leq s(\mathcal{S}(F(x_\alpha^\delta), y) + \mathcal{S}(y, y^\delta)).$$

Combining these inequalities with (15), we obtain

$$\mathcal{S}(F(x_\alpha^\delta), y^\delta) + s \sum_k \alpha_k (\mathcal{R}_k(x_\alpha^\delta) - \mathcal{R}_k(x^\dagger)) \leq s^2 d_{K,y}(F^\delta, F) + (s^2 + s)\mathcal{S}(y, y^\delta).$$

Using the reverse triangle inequality

$$\mathcal{S}(F(x_\alpha^\delta), y^\delta) \geq \frac{1}{s}\mathcal{S}(F(x_\alpha^\delta), y) - \mathcal{S}(y^\delta, y)$$

and the variational inequality (12), we arrive at the estimate

$$\begin{aligned} \frac{1}{s}\mathcal{S}(F(x_\alpha^\delta), y) - s \sum_k \alpha_k \Phi_k(\mathcal{S}(F(x_\alpha^\delta), y)) + s \sum_k \alpha_k D^k(x_\alpha^\delta; x^\dagger) \\ \leq s^2 d_{K,y}(F^\delta, F) + (s^2 + s)\mathcal{S}(y, y^\delta) + \mathcal{S}(y^\delta, y) \end{aligned}$$

Because, by definition of Ψ ,

$$\mathcal{S}(F(x_\alpha^\delta), y) - s^2 \sum_k \alpha_k \Phi_k(\mathcal{S}(F(x_\alpha^\delta), y)) \geq -\Psi(\alpha),$$

we obtain

$$\sum_k \alpha_k D^k(x_\alpha^\delta; x^\dagger) \leq s d_{K,y}(F^\delta, F) + (s+1)\mathcal{S}(y, y^\delta) + \frac{1}{s}\mathcal{S}(y^\delta, y) + \frac{\Psi(\alpha)}{s^2}.$$

Moreover, (14) holds, because every Bregman distance D^j is non-negative and so are the regularisation parameters α_j . \square

Remark 4.3. If one wants to obtain the best possible rate for one specific Bregman distance $D^j(\cdot; x^\dagger)$ with the help of (14), then one should set all regularisation parameters α_i with $i \neq j$ to zero. Indeed, the right hand side of (14) depends on α_i with $i \neq j$ only through Ψ , which is a monotonically increasing function. ■

5 Conclusion

In this paper, we have shown that multi-parameter Tikhonov regularisation is a well-posed regularisation method in the sense of Tikhonov in a fairly general setting on topological spaces with lower semi-continuous and coercive regularisation terms and general distance like measures as similarity term. In addition to proving stability of the regularisation method with respect to noise in the data and varying regularisation parameters, we have also derived the continuous dependence of the regularised solutions on the operator $F: X \rightarrow Y$. To that end, we have introduced a topology on the space of mappings from X to Y that is compatible with the distance measure \mathcal{S} on Y serving as a similarity term. In the classical setting of bounded linear operators between Banach spaces and the squared norm of the residual as a similarity measure, a sequence of bounded linear mappings converges with respect to this topology, if and only if it converges with respect to the norm on $L(X, Y)$.

In addition, we have shown the convergence of the regularised solutions to a true solution, if the regularisation vector decreases to zero slowly enough in dependence of the noise level of the data and the operator error. Extrapolating the results from single-parameter regularisation, one would expect that the limit x^\dagger of the regularised solutions minimises, over the set of all solutions of the equation $F(x) = y$, a certain convex combination $\sum_k \bar{\alpha}_k \mathcal{R}_k$ of the regularisation terms. It seems, however, that this need not be the case if the different components of the regularisation vector converge to zero at different rates. Then one only obtains that

$$\sum_k \bar{\alpha}_k \mathcal{R}_k(x^\dagger) \leq \sum_k \bar{\alpha}_k \mathcal{R}_k(\tilde{x})$$

for every $\tilde{x} \in X$ satisfying $F(\tilde{x}) = y$ and $\mathcal{R}_k(\tilde{x}) < +\infty$ for every k . If one of the coefficients $\bar{\alpha}_k$ equals zero, say $\bar{\alpha}_{k_0} = 0$ this result says nothing about the behaviour of $\sum_k \bar{\alpha}_k \mathcal{R}_k$ on solutions \tilde{x} of the equation $F(x) = y$ satisfying $\mathcal{R}_{k_0}(\tilde{x}) = +\infty$.

Finally, we have derived a quantitative estimate for the difference between the regularised solution and the true solution x^\dagger under the assumption that a variational inequality at the solution x^\dagger is satisfied. Similarly as the stability and convergence result, also this estimate takes into account the effect of operator errors. Also in the case of the quantitative estimates, the difference between multi-parameter regularisation and single-parameter regularisation becomes relevant, as the estimates in the multi-parameter setting cannot be directly translated into optimal convergence rates, as it is not clear, with respect to which distance this optimality should be measured.

References

- [1] G. Aubert and P. Kornprobst. *Mathematical problems in image processing*, volume 147 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2006. Partial differential equations and the calculus of variations, With a foreword by Olivier Faugeras.
- [2] R.I. Boţ and B. Hofmann. An extension of the variational approach for obtaining convergence rates in regularization of nonlinear ill-posed problems. *J. Integral Equations Appl.*, 22:369–392, 2010.
- [3] C. Brezinski, M. Redivo-Zaglia, G. Rodriguez, and S. Seatzu. Multi-parameter regularization techniques for ill-conditioned linear systems. *Numer. Math.*, 94(2):203–228, 2003.
- [4] D. H. Brooks, G. F. Ahmad, R. S. MacLeod, and G. M. Maratos. Inverse electrocardiography by simultaneous imposition of multiple constraints. *IEEE Trans. Biomed. Eng.*, 46(1):3–17, 1999.
- [5] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Probl.*, 20(5):1411–1421, 2004.
- [6] D. Düvelmeyer and B. Hofmann. A multi-parameter regularization approach for estimating parameters in jump diffusion processes. *J. Inverse Ill-Posed Probl.*, 14(9):861–880, 2006.
- [7] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.
- [8] J. Flemming. Theory and examples of variational regularization with non-metric fitting functionals. *J. Inverse Ill-Posed Probl.*, 18(6):677–699, 2010.
- [9] J. Flemming and B. Hofmann. A new approach to source conditions in regularization with general residual term. *Numer. Funct. Anal. Optim.*, 31(3):245–284, 2010.
- [10] M. Grasmair. Generalized Bregman distances and convergence rates for non-convex regularization methods. *Inverse Probl.*, 26(11):115014, 2010.
- [11] M. Grasmair, M. Haltmeier, and O. Scherzer. The residual method for regularizing ill-posed problems. Accepted for publication in *Appl. Math. Comput.*, 2011.
- [12] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Probl.*, 23(3):987–1010, 2007.
- [13] K. Ito, B. Jin, and T. Takeuchi. Multi-parameter Tikhonov regularization. arXiv, math.NA:1102.1173v2, 2011.
- [14] S. Lu and S. V. Pereverzev. Multi-parameter regularization and its numerical realization. *Numer. Math.*, 118(1):1–31, 2011.
- [15] S. Lu, S. V. Pereverzev, Y. Shao, and U. Tautenhahn. Discrepancy curves for multi-parameter regularization. *J. Inverse Ill-Posed Probl.*, 18(6):655–676, 2010.
- [16] Y. Meyer. *Oscillating patterns in image processing and nonlinear evolution equations*, volume 22 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures.
- [17] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42(5):577–685, 1989.
- [18] C. Pöschl. *Tikhonov Regularization with General Residual Term*. PhD thesis, University of Innsbruck, Austria, Innsbruck, October 2008.

- [19] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.
- [20] T. I. Seidman and C. R. Vogel. Well posedness and convergence of some regularisation methods for non-linear ill posed problems. *Inverse Probl.*, 5(2):227–238, 1989.
- [21] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. John Wiley & Sons, Washington, D.C., 1977.
- [22] L. Vese and S. Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. *J. Sci. Comput.*, 19(1–3):553–572, 2003. Special issue in honor of the sixtieth birthday of Stanley Osher.
- [23] P. Xu, Y. Fukuda, and Y. Liu. Multiple parameter regularization: Numerical solutions and applications to the determination of geopotential from precise satellite orbits. *J. Geodesy*, 80(1):17–27, 2006.